Critical Reviews in Biochemistry and Molecular Biology, 42:247–258, 2007

ISSN: 1040-9238 print / 1549-7798 online
DOI: 10.1080/10409230701495631

informa
healthcare

Too Many Mutants with Multiple Mutations

John W. Drake
Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

ABSTRACT It has recently become clear that the classical notion of the random nature of mutation does not hold for the distribution of mutations among genes: most collections of mutants contain more isolates with two or more mutations than predicted by the mutant frequency on the assumption of a random distribution of mutations. Excesses of multiples are seen in a wide range of organisms, including riboviruses, DNA viruses, prokaryotes, yeasts, and higher eukaryotic cell lines and tissues. In addition, such excesses are produced by DNA polymerases *in vitro*. These "multiples" appear to be generated by transient, localized hypermutation rather than by heritable mutator mutations. The components of multiples are sometimes scattered at random and sometimes display an excess of smaller distances between mutations. As yet, almost nothing is known about the mechanisms that generate multiples, but such mutations have the capacity to accelerate those evolutionary pathways that require multiple mutations where the individual mutations are neutral or deleterious. Examples that impinge on human health may include carcinogenesis and the adaptation of microbial pathogens as they move between individual hosts.

KEYWORDS spontaneous mutation, mutational clusters, hypermutation

INTRODUCTION

In 1991, Jacques Ninio argued that microbial populations would contain clones initiated by cells that had experienced some transitory attenuation of replication fidelity and therefore expressed a higher-than-average mutation frequency (Ninio, 1991). At least in bacteria, he reasoned, these accidents would not much affect the average mutation frequency, but they could sharply increase the numbers of mutants bearing two or more mutations. Unlike mutator mutants, which are at a selective disadvantage because their offspring carry an ever increasing load of deleterious mutations, cells exposed to transient phenotypic hypermutation would suffer little disadvantage except from their immediate mutations. This was an unusual paper for *Genetics* because it was rather speculative and did not rest upon a body of new experimental evidence, but I was the editor handling the paper, and I believed that it made an important case for a new view of the mutation process. The paper was well received in the mutagenesis community but slowly sank into the misty mid-regions of history.

Address correspondence to John W. Drake, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC, 27709-2233, USA.E-mail: drake@niehs.nih.gov

In the early months of the current millennium, while analyzing large numbers of mutants generated *in vitro* by the DNA polymerase of bacteriophage RB69, I noticed that many contained two or more well separated mutations in the same reporter molecule, each of which would have produced a mutant phenotype as a single mutation. A quick calculation revealed that there were many more such "multiples" than were expected, had all the mutations been distributed at random among reporter molecules. I put the matter aside for a year or so, but the implanted observation made the same phenomenon capture my attention when perusing other mutational spectra, and soon I began to accumulate a considerable list. This story was eventually submitted for publication with the title, "Too many mutants with multiple mutations" and was rejected immediately on the grounds that the editorial board abhorred alliteration in titles (although puns seemed to be acceptable). Accordingly, I retitled the paper "Clusters of mutations from transient hypermutability" (Drake *et al*., 2005). However, I prefer "multiples" over "clusters" because the latter has a separate meaning in animal genetics, the identical progeny of a single, usually mitotically expanded germ-line mutation.

The phenomenon of "too many multiples" implies that different mutation frequencies exist within different fractions of the subject populations, whether these be organisms, tissues, cells, or polymerase molecules. What processes could fuel such a pattern? In addition to this interesting question, it was also immediately obvious that sufficient frequencies of multiples could drive processes of interest to health professionals. Adaptations requiring more than one new mutation would be accelerated, especially when the single mutations were neutral or deleterious. It is already well known that the adaptation of bacterial pathogens to new hosts can be strongly promoted by heritable mutator mutations and that retroviral infections pass through dramatic mutation-dependent evolutionary phases that recur within serial hosts. It is also clear that carcinogenesis entails a considerable string of mutations, more than would be supported by ordinary rates of single mutations, and might be similarly accelerated not only by mutator mutations (Loeb *et al*., 2003), but also by transient phenotypic hypermutation (but see the constraint described below in the Separations section).

CRITERIA

Early in the process of collecting examples of multiples, whether or not in excess of expectations based on randomness, it became clear that one should distinguish, and remove from consideration, certain kinds of multiples that appeared to be, or probably were, the consequence of a single error. One of these involved clusters of adjacent changes (mostly base-substitution doublets) for which it seemed likely that the first error strongly predisposed to the next. Another occurred among closely spaced clusters that could be described as a single templating event involving, for instance, template switching and imperfect palindromes. In addition, deletions so large as to spare little of the mutation-reporter sequence should be removed from consideration.

In a few systems, it is possible to distinguish between mutations that do and do not produce a mutant phenotype in the reporter system, in which case only the mutations detectable as singles should be considered first, with the piggybacking, singly undetectable mutations considered separately as a special class. Often this distinction cannot be made, but because the undetectable mutations are typically only a few-fold more frequent than the detectable mutations, whereas multiples are often in large excess over the expectations of a random distribution, the conclusion of "too many multiples" is not abrogated.

Most of the estimates of expected numbers of multiples that follow are easily made. If mutations are randomly distributed among reporter sequences, then, by the Poisson distribution, the expected number of multiples depends on M, the number of sequenced mutants, and F, the mutant frequency. The average number of mutations per mutant $a = -\ln(1-F)$. The expected number of mutants with two mutations ("doubles") is $E_2 = Ma^2e^{-a}/2F$. For $F \leq 0.1$, this expression is closely approximated by $E_2 = MF/2$; for instance, for F = 0.1, the longer calculation predicts 0.0996 doubles, the shorter 0.1000. For higher multiples, $E_3 = aE_2/3$ and $E_{i+1} = aE_i/(i+1)$.

UBIQUITY

Table 1 presents a list of spectra with multiples (Drake *et al*., 2005) supplemented with a few spectra accidentally encountered since then. Because I anticipated little further gain from an exhaustive search of

TABLE 1 **"Multiples" in mutational spectra**

System	Genotype or strain	Reporter gene	F	M	D_{exp}	D_{obs}	T_{obs}	$>T_{obs}$	Notes	Reference
Tobacco mosaic virus	WT	*MP*	4.3×10^{-2}	17	0.40	3	3			Malpica *et al*., 2002
HIV-1 RT *in vitro*	WT	*lacZα*	6.4×10^{-2}	434	13.9	24			*a,b*	Bebenek *et al*., 1989
		lacZα	2.3×10^{-2}	99	1.2	2			*a,b*	Eckert and Kunkel, 1993
			1.3×10^{-1}	97	6.2	19	1		*a,b*	Eckert and Kunkel, 1993
Bacteriophage T4	WT	*ac*	1.6×10^{-5}	68	0.00053	0				Wang and Ripley, 1998
	43tsL141	*ac*	1.8×10^{-5}	170	0.0016	0				Wang and Ripley, 1998
T4 pol gp43 *in vitro*	Exo$^-$	*lacZα*	1.1×10^{-2}	121	0.65	2			*b*	Kroutil *et al*., 1998
Phage T4/RB69	WT	*rI*	3.1×10^{-5}	79	0.0012	0			*c*	Bebenek *et al*., 2001
	Exo$^-$	*rI*	2.0×10^{-2}	72	0.72	3			*c*	Bebenek *et al*., 2001
	Pol$^{Y567A/S/T}$	*rI*	2.7×10^{-2}	147	2.0	3			*c*	Bebenek *et al*., 2001
RB69 pol gp43 *in vitro*	Pol^{Y567A}	*lacZα*	2.1×10^{-2}	1324	14	151	6		*b,d*	Drake *et al*., 2005
		lacZα^{ud}	1.9×10^{-1}	237	22	23	1		*b,e*	Drake *et al*., 2005
Herpes simplex virus	WT	*supF*	4.9×10^{-4}	80	0.020	7	1		*f*	Hwang *et al*., 1999
			1.3×10^{-4}	85	0.0054	0			*f*	Hwang *et al*., 2002
		tk	6.0×10^{-5}	66	0.0020	1			*f*	Lu *et al*., 2002
	PAAr5	*supF*	1.0×10^{-3}	87	0.045	4	2		*f*	Hwang *et al*., 1999
	Y7	*supF*	1.9×10^{-3}	53	0.050	0	1		*f*	Lu *et al*., 2002
			4.7×10^{-4}	92	0.021	0			*f*	Hwang *et al*., 2002
		tk	4.0×10^{-2}	66	1.3	6			*f*	Lu *et al*., 2002
	Y7 Exo$^-$	*supF*	4.8×10^{-3}	249	0.60	11			*f*	Hwang and Hwang, 2003
	YD12	*supF*	1.5×10^{-3}	77	0.059	2			*f*	Hwang *et al*., 1999
Escherichia coli	WT	*supF*	2.1×10^{-7}	38	0.000004	1			*f*	Akasaka *et al*., 1992
		lacI	2.0×10^{-6}	167	0.00017	0				Schaaper *et al*., 1986
		lacI^d	1.3×10^{-7}	368	0.00002	2				Oller and Schaaper, 1994
		lacI^d	1.1×10^{-7}	413	0.00002	0				Schaaper and Dunn, 1991
			2.4×10^{-7}	269	0.00003	0				Sargentini and Smith, 1994
		egbR	3.0×10^{-8}	73	0.000001	1				Hall, 1999
	mutD5	*lacI*^{d}	1.5×10^{-3}	498	0.37	4				Schaaper, 1988
	mutL$^-$	*lacI*^{d}	3.5×10^{-5}	243	0.0043	2				Schaaper, 1993
		lacI^d	1.2×10^{-5}	196	0.0011	1				Schaaper, 1993
	mutHLS	*lacI*^{d}	2.6×10^{-7}	487	0.00063	0				Schaaper and Dunn, 1987
	dnaE911	*lacI*^{d}	0.8×10^{-7}	476	0.00002	1				Oller and Schaaper, 1994

(Continued on next page)

TABLE 1 **"Multiples" in mutational spectra *(Continued)***

System	Genotype or strain	Reporter gene	F	M	D_{exp}	D_{obs}	T_{obs}	$>T_{obs}$	Notes	Reference
	dnaE173	*rpsL*	9.2×10^{-6}	56	0.00026	1	1			Mo *et al*., 1991
E. coli pol I(K) *in vitro*	WT	*lacZα*	4.7×10^{-3}	118	0.28	3			*b,g*	Bell *et al*., 1997
	Y766A/S	*lacZα*	3.8×10^{-2}	224	4.3	5			*b,g*	Bell *et al*., 1997
Sulfolobus acidocaldarius	WT	*pyrE*	3.4×10^{-7}	108	0.00002	0				Grogan *et al*., 2001
Saccharomyces cerevisiae	WT	*SUP4*-o	1.9×10^{-6}	297	0.00028	2				Kunz *et al*., 1990
	rad1	*SUP4*-o	1.3×10^{-5}	242	0.0015	1				Kunz *et al*., 1990
		URA3	2.2×10^{-6}	35	0.00004	0				Lee *et al*., 1988
	WT	*CAN1*	2.0×10^{-7}	21	0.000002	2			*h*	Venkatesan *et al*., 2006
	pol3$^{L612F/M/K/G}$	*CAN1*	2.0×10^{-6}	76	0.000075	3			*h*	Venkatesan *et al*., 2006
S. cerevisiae pol *δ in vitro*	WT	*lacZα*	3.3×10^{-3}	182	0.30	1			*b*	Nick McElhinny *et al*., 2007
	L612M	*lacZα*	2.2×10^{-2}	401	4.3	13			*b*	Nick McElhinny *et al*., 2007
	D520V	*lacZα*	2.3×10^{-2}	486	5.7	9			*b*	Nick McElhinny *et al*., 2007
Rat cell line	WT	*cII*	1.3×10^{-4}	99	0.0064	1			*i*	Watson *et al*., 1988
Mouse cell line	WT	*gpt*	2.0×10^{-5}	43	0.00043	0		1(5)	*j*	Ashman and Davidson, 1987
Chinese hamster cell line	WT	*gpt*	1.2×10^{-4}	18	0.0011	2			*k*	Romac *et al*., 1989
			1.3×10^{-5}	58	0.00038	0			*k*	Tindall and Stankowski, 1989
Monkey cell line	WT	*supF*	8.2×10^{-4}	120	0.049	0			*f*	Cabral-Neto *et al*., 1993
Human cell lines	WT	*HPRT*	9.0×10^{-6}	200	0.0009	6	1		*l*	Lichtenauer-Kaligis *et al*., 1996
			1.8×10^{-5}	51	0.00046	0				Ikehata *et al*., 1989
			3.0×10^{-6}	33	0.00005	0				Giver *et al*., 1993
Mouse tissue	WT	*cII*	9.5×10^{-5}	182	0.0086	1			*m*	Harbach *et al*., 1999
		lacI	4.2×10^{-5}	348	0.0073	2			*n*	de Boer *et al*., 1997
			2.3×10^{-5}	435	0.0050	7		1(5)	*o*	Buettner *et al*., 2000
Monkey tissue	WT	*HPRT*	3.0×10^{-6}	40	0.00006	0			*p*	Harbach *et al*., 1995
Human tissue	WT	*HPRT*	1.9×10^{-4}	82	0.0078	5		1(4)		Colgin *et al*., 2002
			6.0×10^{-6}	31	0.00009	0				Rossi *et al*., 1990
Rat hepatoma pol *β in vitro*	WT	*lacZα*	1.1×10^{-1}	296	15.7	≤16			*b,q*	Kunkel, 1985

TABLE 1 **"Multiples" in mutational spectra *(Continued)***

System	Genotype or strain	Reporter gene	F	M	D_{exp}	D_{obs}	T_{obs}	$>T_{obs}$	Notes	Reference
Chick embryo pol β *in vitro*	WT	*lacZα*	7.3×10^{-2}	144	5.2	$\leq$1			[b,q]	Kunkel, 1985
Rat pol β* *in vitro*	WT	*HSV-tk*	1.4×10^{-3}	86	0.060	2			[r]	Opresko *et al*., 1998
	T79S	*HSV-tk*	2.7×10^{-3}	79	0.11	3	3	6(4–9)	[r]	Maitra *et al*., 2002
	Y265C	*HSV-tk*	4.4×10^{-2}	79	1.7	31	8	6(4–5)	[r]	Maitra *et al*., 2002

All of the Systems are *in vivo* except for the several DNA polymerases, which are labeled "*in vitro.*" The reporter gene is natural unless noted to be a transgene. WT = wild type. F = frequency of spontaneous mutants, adjusted where possible and appropriate for the efficiency of detecting mutants and for mutations that do not produce a phenotype. M = number of mutants sequenced exclusive of large deletions and ignoring insertions of mobile elements. D_{exp} = number of doubles expected from a random distribution of mutations. D_{obs} = observed number of mutants with two mutations, T_{obs} = observed number of mutants with three mutations, $>T_o$ = observed number of mutants with more than three mutations, with numbers of mutations inside (s). Multiple mutations exclude tandem mutations, synonymous mutations, and complex mutations that arise repeatedly and appear to be templated by a specific, imperfect and usually reverse repeat.

[a]Human immunodeficiency virus reverse transcriptase.

[b]About 40% of *lacZα* mutants are lost during the assay, so that F = (observed mutant frequency)/0.6. The system distinguishes between mutations that are detectable or not when present as singles, and only detectable mutations are tabulated.

[c]The RB69 and T4 replicases (gp43) have both polymerase (Pol) and exonuclease proofreading (Exo) sites. The data derive from experiments in which T4 replication was driven by a plasmid-borne RB69 gp43 and the *rI* reporter gene was in T4. The Pol$^{Y567A/S/T}$ entry is the sum of three different substitutions at the RB69 gp43 Y567 with very similar mutator properties, F being the combined value for each mutant weighted by its M.

[e]The data derive from the same experiments as above but the mutations were undetectable as singles but were detected as piggybackers on detectable mutations. $F = 1.851 \times 10^{-1}$ was the combined value for each supplement regimen weighted by its M. The expected number of triples was 1.5.

[f]*supF* is a tRNA transgene from *E. coli* and of a type that may be generally hypermutable, whereas *tk* is an endogenous gene and displays an approximately normal mutation rate.

[g]K indicates the Klenow fragment of pol I. Y766A/S are mutator mutants.

[h]The data derive from fluctuation tests for which F values were not available, but in such tests, the mutation rate and F are similar, so the former were used. The *pol3* data were pooled from four mutator mutants, F being the weighted average of the individual values.

[i]Embryonic fibroblast cell line with *cII* transgene from phage λ.

[j]A9 cell line with *gpt* transgene from *E. coli*. The 5 mutations were BPSs scattered throughout *gpt*.

[k]Ovary cell line with *gpt* transgene from *E. coli*. First report also listed one double containing one missense and one synonymous mutation.

[l]TK6 lymphoblastoid cell line with a human *HPRT* cDNA transgene at five different sites. Also two doubles containing one missense or indel and one synonymous mutation.

[m]Mouse liver, lung and spleen with *cII* transgene from phage λ.

[n]Mouse liver with *lacI* transgene from *E. coli*.

[o]Numerous mouse tissues with a *lacI* transgene from *E. coli*.

[p]*HPRT* mutations arising *in vivo* in cynomolgus monkey T-lymphocytes.

[q]Because doubles were combined with duplications and complex mutations, the D values are $\leq$.

[r]Recombinant enzyme made in *E. coli* and bearing an added Gly-Ser at its 5′ end.

all published spectra, the list is necessarily somewhat anecdotal, which does not matter in the present context and may provide a pleasing exercise to someone else. Table 1 also lists the spectra in the original collection that lacked examples of multiples, for which display space was previously not adequate. The list is ordered by genomic complexity from riboviruses to humans, with the examples of polymerases examined *in vitro* directly following the corresponding organism. The current list contains 39 examples of multiples in excess, in a few cases by small factors and in many by large factors. The list also contains 17 examples where no multiples were observed. Finally, the list contains seven examples where multiples were observed but were not in excess of the predictions of randomness; some of these are revealing and will be considered later.

Almost all organisms for which mutational spectra are available, and almost all similarly tested DNA polymerases, produce more multiples than predicted from random distributions in at least one spectrum (Drake *et al.*, 2005). Many such spectra contain only one or two multiples. Therefore, many of the spectra displaying no multiples would probably display one or more had many more mutants been examined.

The kinds and relative frequencies of mutations observed in multiples, including base- or base-pair substitutions (BPSs) or indels (insertions or deletions of any size but most commonly single-base deletions and insertions), are usually similar to those observed in mutants with single mutations (data not shown).

TRIVIALITIES

A key step in obtaining the DNA sequences of mutation reporters is often amplification by the polymerase chain reaction, which is usually done with a somewhat error-prone enzyme and is sometimes preceded by much more error-prone reverse transcription. Provided that sufficient RNA or DNA molecules are present at the first step of amplification, the introduced errors will be a small fraction of the final molecules and will not provide false multiples. All of the entries in Table 1 meet this criterion.

It is instructive to compare certain parameters of spectra with multiples in excess versus spectra with no multiples at all (Table 2).

One trivial explanation for multiples is that they are sequencing errors (to be considered separately from amplification errors); that is, they do not exist except as a result of the act of measurement (somewhat à la Heisenberg). Sequencing errors presumably do occur. If they were sufficiently frequent to impact the overall pattern, what parameter of Table 2 might reveal this? Taking an analytical hint from microeconomics (Levitt and Dubner, 2006), one may surmise that the frequency of sequencing errors will have decreased during the steady procedural improvements of the past two decades, so that the proportion of spectra with an excess of multiples would decrease over time. However, it does not (Table 3) but instead rises. In addition, the frequency of spectra with multiples does not appear to vary inversely with the quality of the research group (data not shown). Thus, sequencing errors are unlikely to create an important fraction of observed multiples.

Two other parameters of Table 2, *M* and *F*, might relate to the frequency with which spectra display multiples. The impacts of these factors are shown in Table 4. First, the more mutants sequenced, the greater should be the chance of encountering an infrequent multiple. The numbers fall in the expected direction but with unconvincing *p* values, so that, within the range of these examples, *M* has at most a weak effect on the detection of multiples. Second, because the proportions of *expected* multiples in a spectrum must increase with *F*, it will be correspondingly easier to detect an excess of multiples. This appears to be a strong determinant of detection. Note also that there is no obvious *a priori* reason why the frequency of sequencing errors should vary with the observed mutant frequency.

TABLE 2 **Spectra with numbers of multiples in excess of expectations from random distributions (+) or with no multiples at all (–) as functions of mutant frequency (*F*), number of mutants sequenced (*M*), and year of publication (*Y*)***

Multiples	*F*	–log*F*	*M*	*Y*
+	3.0×10^{-8}	7.52	73	1999
+	8.0×10^{-8}	7.10	476	1994
+	1.3×10^{-7}	6.89	368	1994
+	2.0×10^{-7}	6.70	21	2006
+	2.1×10^{-7}	6.68	38	1992
+	1.9×10^{-6}	5.92	297	1990
+	2.0×10^{-6}	5.70	76	2006
+	9.0×10^{-6}	5.05	200	1996
+	9.2×10^{-6}	5.04	56	1991
+	1.2×10^{-5}	4.92	196	1993
+	1.3×10^{-5}	4.89	242	1990
+	2.0×10^{-5}	4.70	43	1987
+	2.3×10^{-5}	4.64	435	2000
+	3.5×10^{-5}	4.56	243	1993
+	4.2×10^{-5}	4.38	348	1997
+	6.0×10^{-5}	4.22	66	2002
+	9.5×10^{-5}	4.02	182	1999
+	1.2×10^{-4}	3.92	18	1989
+	1.3×10^{-4}	3.88	99	1988
+	1.9×10^{-4}	3.72	82	2002
+	4.9×10^{-4}	3.31	80	1999
+	1.0×10^{-3}	3.00	87	1999
+	1.4×10^{-3}	2.85	86	1998
+	1.5×10^{-3}	2.82	77	1999
+	1.5×10^{-3}	2.82	498	1988
+	1.9×10^{-3}	2.72	53	2002
+	2.7×10^{-3}	2.57	79	2002
+	3.3×10^{-3}	2.48	182	2007
+	4.7×10^{-3}	2.33	118	1997
+	4.8×10^{-3}	2.32	249	2003
+	1.1×10^{-2}	1.96	121	1998
+	2.0×10^{-2}	1.70	72	2001
+	2.1×10^{-2}	1.68	1324	2005
+	2.2×10^{-2}	1.66	401	2007
+	4.0×10^{-2}	1.40	66	2002
+	4.4×10^{-2}	1.36	79	2002
+	4.7×10^{-2}	1.33	17	2002
+	6.4×10^{-2}	1.19	434	1989
+	1.3×10^{-1}	0.89	97	1993
–	1.1×10^{-7}	6.96	413	1991
–	2.4×10^{-7}	6.62	269	1994
–	2.6×10^{-7}	6.59	487	1987
–	3.4×10^{-7}	6.47	108	2001
–	2.0×10^{-6}	5.70	167	1986
–	2.2×10^{-6}	5.66	35	1988
–	3.0×10^{-6}	5.52	33	1993
–	3.0×10^{-6}	5.52	40	1995
–	6.0×10^{-6}	5.22	31	1990
–	1.3×10^{-5}	4.89	58	1989
–	1.6×10^{-5}	4.80	68	1998

TABLE 2 **Spectra with numbers of multiples in excess of expectations from random distributions (+) or with no multiples at all (–) as functions of mutant frequency (*F*), number of mutants sequenced (*M*), and year of publication (*Y*)* *(Continued)***

Multiples	*F*	–log*F*	*M*	*Y*
–	1.8×10^{-5}	4.74	51	1989
–	1.8×10^{-5}	4.74	170	1998
–	3.1×10^{-5}	4.51	79	2001
–	1.3×10^{-4}	3.89	85	2002
–	4.7×10^{-4}	3.33	92	2002
–	8.2×10^{-4}	3.09	120	1993

*The table omits seven entries from Table 1 in which multiples were present in nearly the predicted numbers: HIV-1 RT *in vitro*, second entry; phage T4/RB69 *in vitro*, third entry; RB69 pol gp43 *in vitro*, second entry; *E. coli* pol I(K) *in vitro*, second entry; *S. cerevisiae* pol δ *in vitro*, third entry; and both rat hepatoma and chick embryo pol β *in vitro*.

If the mutant frequency is underestimated, then the expected frequency of multiples will be underestimated. Why might *F* be underestimated? No mutation screen detects all mutations; for instance, most base-substitution mutations produce a phenotype too weak to be detected under all laboratory screens except blind sequencing. However, because the analysis is usually restricted to *detectable* mutations, the calculations hold; as noted previously, undetectable mutations piggybacking upon detectables are usually only a few-fold more frequent than multiples composed exclusively of detectable mutations, whereas factors by which multiples are in excess are usually larger.

TABLE 3 **Spectra with multiples in excess (Yes) or absent (No) as a function of year of publication**

	Multiples		Fraction with multiples in excess
Year	Yes	No	
1986	0	1	
1987	1	1	
1988	2	1	
1989	2	2	
1990	2	1	
1991	1	1	
1992	1	0	
1993	3	2	
1994	2	1	
1995	0	1	
1996	1	0	
1997	2	0	
1998	2	2	
1999	5	0	
2000	1	0	
2001	1	2	
2002	7	2	
2003	1	0	
2004	0	0	
2005	1	0	
2006	2	0	
2007	2	0	
86–90	7	6	0.54
91–95	7	5	0.58
96–00	11	2	0.85
01–05	10	4	0.71
86–96	15	11	0.58
97–07	24	6	0.80

A more interesting contributor to false excesses might be phenotypic lag, the interval between the creation of a mutation and the time its expression becomes sufficient for phenotypic detection. If the lag lasts for several generations, which is sometimes the case, then both *F* and thus the predicted number of multiples will be underestimated, and excesses of multiples will be correspondingly overestimated. The effect can be specified. Putting aside for present purposes the stochastic nature of mutation, as a population grows, it begins to mutate as its size approaches the reciprocal of the mutation rate (although the first mutation will usually have arisen at some earlier time). In that "first" mutating generation ($g \equiv 1$), on average, one mutation occurs, producing one mutant and one nonmutant progeny. (If one mutation produces two mutant progeny [Witkin and Sicurella, 1964; Witkin and Parisi, 1974], the argument must be recast but the overall results will be similar.) The mutant goes on to generate a clone of $2^{(g-1)}$ mutants in the succeeding $(g-1)$ generations until the mutants are scored. In the next generation, two mutations occur on average, and each generates a clone of size $2(g-2)$, thus contributing $2 \times 2^{(g-2)} = 2^{(g-1)}$ mutants to the final population. Thus, assuming that mutant and non-mutant cells grow at the same rate, each mutating generation contributes the same number of mutant cells to the eventual total of $g2^{(g-1)}$ mutants. If phenotypic lag lasts for p generations, then the total number of detected (= expressed) mutants will be $(g-p)2^{(g-1)}$. Thus, the efficiency of detection will be the ratio of expressed mutants to total mutants, $(g-p)/g$, and the effect of phenotypic lag will decrease as g increases. In practice, most mutation-reporter systems have been adopted in part because they express mutations well, which includes having short phenotypic lags (such as a single generation). An interesting exception may be the halophilic archaeon *Haloferax volcanii*, in which the number of chromosomes per cell appears to be very large so that considerable segregation must occur until

TABLE 4 Parameters of mutants with and without excess multiples

Group	No. of spectra	Mean *M*	Median *M*	Median –log*F*	Median *F*
Excess multiples	39	196.9	97	3.72	1.9×10^{-4}
No multiples	17	135.6	85	5.22	6.0×10^{-6}
p for difference*		0.31	0.77	0.0015	

*The *p* values are two-sided, for the mean *M* by the Mann-Whitney test, and for the medians by the median test.

a recessive mutation can be expressed (Mackwan *et al*., 2007).

Multiples can arise simply by sequential accumulations of singles rather than in virtually simultaneous bursts, but if all the genomes share a continuously constant mutation rate, then multiples will not occur in excess.

If most multiples were caused by mutator mutants in the population, then the phenomenon would not be particularly novel. A semiquantitative argument was made previously (Drake *et al*., 2005) that took into account the frequency of mutators in cultures of laboratory-adapted *E. coli* and *Salmonella typhimurium* (lower than expected from mutation pressure, presumably because of the reduced fitness of mutators in the absence of strong selection for new mutations) and the impact of strong bacterial mutators on the average gene. The observed frequency of doubles F_D is the mutant frequency F times the proportion of doubles among mutants ($F_D = FD/M$). Strong mutator mutations increase the value of F for a gene by at most about 100-fold (stimulating mutation at some sites by much more but at many sites by much less). In laboratory populations, the frequency of mutator mutants is $\leq 10^{-5}$. Thus, the frequency of doubles due to mutators will be $\leq (100\text{F})^2(10^{-5}) = F^2/10$ and the fraction of doubles caused by mutators will be $\leq FM/10D$. This fraction is ≤ 0.1 for all of the entries in Table 1 for which D_{obs} is substantially greater than D_{exp}, so that in such bacteria, at least, heritable mutator mutations produce few of the observed doubles.

RATES

Mutation rates are usually estimated from mutant frequencies, population sizes (a measure of the number of generations), and key assumptions about the geometry of replication (such as exponential, linear, or mixed). In most organisms whose genomes are encoded by DNA rather than RNA, replication is exponential and the mutation rate μ is assumed to be the same for all replication events. This cannot apply for transient hypermutation during replication, in which multiples may arise within a single generation but propagate with the standard μ thereafter. However, there are at least two methods to fractionate a population into two or more subpopulations, each with its own fractional composition of the population as a whole and its own mutation frequency. Given sufficient sample sizes, it might then become possible to model the distribution of hypermutagenic events during the growth of the total population, although such modeling will not be presented here.

One method, whose derivation was provided previously (see Supporting Text in Drake *et al*., 2005), depends on the availability of a spectrum with both doubles (D) and triples (T), and on the condition that most of the multiples were contributed by the minority, high-F subpopulation. Assuming that a population contains two subpopulations, S_1 and S_2 ($S_1 > S_2$), with respective mutation frequencies F_1 and F_2 ($F_1 < F_2$), then $F_2 = 3T_{obs}/D_{obs}$, $S_2 = 2FD_{obs}/MF_2{}^2\text{e}^{-F2}$, $S_1 = 1-S_2$, and $F_1 = -\ln[(1-F-S_2\text{e}^{-}F_2)/S_1]$. This method is strikingly simple but would be difficult to extend to more than two subpopulations because of the difficulty of untangling multiples produced by more than one hypermutating subpopulation.

Another method is based on the progressive fitting of the four parameters to the data using mixtures of Poisson distributions and an iterative expectation-maximization procedure. This method can be expanded to more than two subpopulations but quickly becomes calculation-intensive with additional subpopulations and/or increasing M.

When these methods were applied to the TMV results summarized in Table 1, they produced similar results (Table 5): most of the population had $F \approx 0.01$ (instead of the observed $F \approx 0.04$) while 3 to 4% of the population had $F = 1$ to 3%. Note that the iterative method can be continued until a very large p

TABLE 5 **Deconstructing a TMV population**

Method	S_1	F_1	S_2	F_2	p
Algebraic	0.97	0.011	0.03	3.0	0.40
Iterative	0.96	0.013	0.04	1.2	0.98

The TMV population had $F = 4.253 \times 10^{-2}$, $M = 17$, $D_{exp} = 0.40$, $D_{obs} = 3$, $T_{exp} = 0.0058$, and $T_{obs} = 3$.

value is obtained, as here, and can be supplemented by additional subpopulations if desired.

In practice, because transient hypermutability can be imagined to occur by diverse mechanisms (as discussed later), it is reasonable to surmise that populations often contain more than two characteristic S_i/F_i fractions. Indeed, a population of molecules produced by the DNA polymerase of phage RB69 simply could not be analyzed by the above methods. However, the methods may be helpful when most of the multiples are produced by only one subpopulation.

SEPARATIONS

Most mutation reporters are 10^2 to 10^3 bases long. Most spectra have too few multiples to discern whether they are clustered or scattered within the reporter sequence. In a few cases, however, the distance distribution between the components of multiples has been determined.

Among mutants of the human *HPRT* gene in normal primary kidney tubular epithelial cells, multiples were in strong excess (Colgin *et al*., 2002): $F \approx 2 \times 10^{-4}$, $M = 82$, $D_{exp} \approx 0.008$, $D_{obs} = 5$. Ignoring one pair of adjacent mutations, the distances between doubles were 1, 6, 5012, 7023, and 25024 intervening bases. In addition, one quadruple was observed, with distances of 13, 214 and 4895 intervening bases. These distances clearly display strongly nonrandom clustering ($p \approx 0.002$), suggesting "a mechanistically linked origin." In contrast, multiples whose components arose sequentially rather than in a burst had an intervening-base distribution expected for mutations of independent origin ($p \approx 0.8$) (Finette *et al*., 2000).

A nonrandom pattern of clustering was also observed among multiples arising in an *E. coli* transgenic *lacI* reporter in the mouse (Hill *et al*., 2004), where the separations were exponentially distributed with a median separation of 120 bases in a target of about 1.4 kilobases (kb), and were thus described as "chronocoordinate" events. Subsequently, 65 such multiples were subjected to resequencing over about 20 kb outside of the reporter sequence, and ten were found to contain additional mutations (Wang *et al*., 2007). (When 130 singles were thusly sequenced, only one had a single mutation in the outside regions.) When these additional mutations are considered, the size of the hypermutated region turned out to be $\leq$30 kb.

Multiples produced *in vitro* by the DNA polymerase of phage RB69, however, were composed of a random sample of all mutations (Drake *et al*., 2005). Although the RB69 target was smaller ($\approx$270 bases) than for the mammalian studies performed *in vivo* ($\approx$1000–25000 bases), the small-distance tail of the mammalian distributions encompassed the RB69 distribution, so the difference is probably real.

These results reveal that the tract of transient hypermutation can extend from a few bases to about 30 kb, and that the shapes of the distribution of intervening bases depends on the system. A further insight can be gained from the TMV results (Malpica *et al*., 2002). With a reporter sequence of 804 bases, three doubles had separations of 210, 236 and 313 intervening bases and three triples had separations in their component doubles of 21, 279, 286, 336, 336, and 641 intervening bases. Here there is no hint of clustering. However, if the mutation frequency in the reporter sequence applied to the entire 6395-base genome, then these 6 mutants would have had an average of roughly 20 mutations per genome. About two thirds of these mutations would have been indels, and the average base substitution is probably more deleterious in riboviruses than elsewhere, so these genomes would not have survived. Therefore, either transient hypermutation in this system is spatially delimited despite the lack of clustering, or only the tail of a long distribution could be observed among viable multiples whose mutations happened to fall exclusively within the reporter gene.

Because the several mutations required for carcinogenesis must hit genes widely scattered in the human genome, the limit of about 30 kb over which multiples were scattered in the above two mammalian examples somewhat decreases the chance that transient phenotypic hypermutation could promote cancer. Alternatively, in the singles and in multiples that showed no outside mutations, a different pattern might lurk, one that scattered multiples far more widely. An entirely new kind of search is needed to address this possibility.

ORIGINS

The causes of transient hypermutation remain obscure. Ninio (1991) pointed out that it could result when errors of transcription or translation produced mutagenic proteins, and mutator DNA polymerases come immediately to mind. Random DNA damage seems to activate SOS systems in a small fraction of bacteria per generation, increasing the mutation rate for perhaps an hour (*e.g.*, Little, 1991; McCool *et al*., 2004). DNA resynthesis following DNA mismatch repair might produce clusters within roughly a kilobase if the new DNA strand was not itself subject to mismatch repair, a matter that remains open (Ninio, 2000), and similar clustering might be produced by certain types of gene-conversion events (Ninio, 1996). An error-prone component in the repair of DNA double-strand breaks might also produce clustered multiples (Ponder *et al*., 2005). When a protein crucial for fidelity is present in small numbers per cell, random asymmetries in its segregation at cell division could yield a transiently hypermutating progeny cell, an example being the MutL and/or MutS components of DNA mismatch repair (Feng *et al*., 1996). Random physiological states that reduce fidelity are also candidates. Protein misfolding is a common anomaly that might contribute to replication errors. Perhaps occasional mutagenic tracts are initiated by an incorrectly positioned or composed replication complex that persists for a while. There is a considerable literature demonstrating that nutrient-deprived stationary-phase cells tend to enter a phenotypically mutable condition, sometimes with a hypermutating subpopulation (Bull *et al*., 2000; Foster, 2004; Ponder *et al*., 2005). Altogether, there is an impressive number of ways in which the best-evolved fidelity schemes gang aft agley (Burns, 1786).

There are obvious strategies to explore some of these possibilities. Errors of translation are sometimes mutagenic and can be reduced by specific chemical or genetic interventions. SOS systems can be repressed mutationally. Chaperones assist protein folding and can be overexpressed, and some proteins can be unfolded and refolded *in vitro* with recovery of function. Cell size can be enhanced, perhaps ameliorating asymmetries of protein segregation. Note that only changes that reduce mutagenesis, rather than enhancing it, are truly informative in such a situation. A manipulation that enhances the frequency of multiples in a particular system may or may not be enhancing a particular mechanism already generating multiples, whereas a manipulation that decreases the frequency of multiples must necessarily inhibit the mechanism that already generates them.

A paradox lurks among the entries of Table 1 for DNA polymerases acting *in vitro*. In order to observe too many multiples, hypermutable synthesis must occur in tracts. However, two of the polymerases that produce excess multiples have limited processivity, substantially shorter than the gaps used as mutation reporters. The processivity of the RB69 polymerase is only about 10^1 to 10^2 consecutive insertions before the enzyme dissociates from its substrate, so that a single molecule will infrequently traverse a 266-base *lacZ*α gap, whereas the number of intervening bases between the components of the numerous observed doubles was not biased toward closeness. Nevertheless, multiples were observed in excess (expected doubles = 14, observed doubles = 151, and 6 triples) (Table 1), and the presence or absence of accessory proteins had little or no effect on the frequencies of multiples (Drake *et al*., 2005). The processivity of pol β is ≈ 1, so no excess of multiples should be observed. That may have been the case for native rat pol β purified in the early 1980s (expected doubles ≈ 16, observed doubles ≤ 16) (Kunkel, 1985), but multiples were in excess with the same but now aged enzyme preparation in 2006 (expected ≈ 9, observed = 18 and three triples) (L. García-Villada and J. W. Drake, unpublished results). With recombinant rat pol β in the late 1990s, two groups reported excess doubles, 0.06 expected versus 2 observed (Opresko *et al*., 1998) and 1.3 expected versus 7 observed (Osheroff *et al*., 1999), although somewhat fewer were detected in yet another test, 3.6 expected versus 6 observed (L. García-Villada and J. W. Drake, unpublished results). With recombinant human pol β, excess doubles were also observed, 1.2 expected versus 6 observed (Osheroff *et al*., 1999). The possibility of a contaminating polymerase in the recombinant rat enzyme seems to be reduced by the observation that a version of the enzyme with its polymerase activity obliterated mutationally failed either to increase the background mutant frequency of the test system significantly or to introduce any multiples (L. García-Villada and J. W. Drake, unpublished results). A resolution of the apparent paradox of multiples without processivity is likely to be informative.

ACKNOWLEDGMENTS

I thank Grace Kissling for statistical assistance and Jim Mason and Marilyn Diaz for critical readings of the manuscript. This research was supported by the Intramural Research Program of the NIH, and NIEHS.

REFERENCES

Akasaka, S., Takimoto, K., and Yamamoto, K. 1992. G:C →T:A and G:C →C:G transversions are the predominant spontaneous mutations in the *Escherichia coli supF* gene: an improved *lacZ(am) E. coli* host designed for assaying pZ189 *supF* mutational specificity. *Mol Gen Genet* 235:173–178.

Ashman, C.R. and Davidson, R.L. 1987. Sequence analysis of spontaneous mutations in a shuttle vector gene integrated into mammalian chromosomal DNA. *Proc Natl Acad Sci USA* 84:3354–3458.

Bebenek, A., Dressman, H.K., Carver, G.T., Ng, S., Petrov, V., Yang, G., Konigsberg, W.H., Karam, J.D., and Drake, J.W. 2001. Interacting fidelity defects in the replicative DNA polymerase of bacteriophage RB69. *J Biol Chem* 276:10387–10397.

Bebenek, K., Roberts, J.D., Wilson, S.H., and Kunkel, T.A. 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J Biol Chem* 264:16948–16956.

Bell, J.B., Eckert, K.A., Joyce, C.M., and Kunkel, T.A. 1997. Base miscoding and strand misalignment errors by mutator Klenow polymerase with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J Biol Chem* 272:7345–7351.

Buettner, V.L., Hill, K.A., Scaringe, W.A., and Sommer, S.S. 2000. Evidence that proximal multiple mutations in Big Blue® transgenic mice are dependent events. *Mutat Res* 452:219–229.

Bull, H.J., McKenzie, G.J., and Rosenberg, S.M. 2000. Evidence that stationary-phase hypermutation in the *Escherichia coli* chromosome is promoted by recombination. *Genetics* 154:1427–1437.

Burns, R. 1786. To a mouse, on turning her up in her nest with the plough. In: *Poems, Chiefly in the Scottish Dialect*. Kilmarnock: John Wilson, 138–140.

Cabral-Neto, J.B., Gentil, A., Cabral, R.E.C., and Sarasin, A. 1993. Implication of uracil in spontaneous mutation on a single-stranded shuttle vector replicated in mammalian cells. *Mutat Res* 288:249–255.

Colgin, L.M., Hackmann, A.F.M., and Monnat, Jr., R.J. 2002. The unexpected landscape of *in vivo* somatic mutations in a human epithelial cell lineage. *Proc Natl Acad Sci USA* 99:1437–1442.

de Boer, J.G., Erfle, H., Walsh, D., Holcroft, J., Provost, J.S., Rogers, B., Tindall, K.R., and Glickman, B.W. 1997. Spectrum of spontaneous mutations in liver tissue of *lacI* transgenic mice. *Environ Mol Mutagen* 30:273–286.

Drake, J.W., Bebenek, A., Kissling, G.E., and S. Peddada, S. 2005. Clusters of mutations from transient hypermutability. *Proc Natl Acad Sci USA* 102:12849–12854.

Eckert, K.A. and Kunkel, T.A. 1993. Fidelity of DNA synthesis catalyzed by human DNA polymerase a and HIV-1 reverse transcriptase: effect of reaction pH. *Nucl Acids Res* 21:5212–5220.

Feng, G., Tsui, T.H.-C., and Winkler, M.E. 1996. Depletion of the cellular amount of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase *Escherichia coli* K-12 cells. *J Bacteriol* 178:2388–2396.

Finette, B.A., Homans, A.C., and Albertini, R.J. 2000. Emergence of genetic instability in children treated for leukemia. *Science* 288:514–517.

Foster, P.L. 2004. Adaptive mutation in *Escherichia coli*. *J Bacteriol* 186:4846–4852.

Giver, C.R., Nelson, S.L., and Grosovsky, A.J. 1993. Spectrum of spontaneous HPRT– mutations in TK6 human lymphoblasts. *Environ Mol Mutagen* 22:138–146.

Grogan, D.W., Carver, G.T., and Drake, J.W. 2001. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci USA* 98:7928–7933.

Hall, B.G. 1999. Spectra of spontaneous growth-dependent and adaptive mutations at *ebgR*. *J Bacteriol* 181:1149–1155.

Harbach, P.R., Mattano, S.S., Zimmer, D.M., Wang, Y., and Aaron, C.S. 1995. DNA sequence analysis of spontaneous *hprt* mutations arising in vivo in cynomolgus monkey T-lymphocytes. *Environ Mol Mutagen* 26:218–225.

Harbach, P.R., Zimmer, D.M., Filipunas, A.L., Mattes, W.B., and Aaron, C.S. 1999. Spontaneous mutation spectrum at the lambda *cII* locus in liver, lung, and spleen tissue of Big Blue transgenic mice. *Environ Mol Mutagen* 33:132–143.

Hill, K.A., Wang, J., Farwell, K.D., Scaringe, W.A., and Sommer, S.S. 2004. Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency. *Mutat Res* 554:223–240.

Hwang, Y.T. and Hwang, C.B.C. 2003. Exonuclease-deficient polymerase mutant of herpes simplex virus type 1 induces altered spectra of mutations. *J Virol* 77:2946–2955.

Hwang, Y.T., Liu, B.Y., Hong, C.Y., Shillitoe, E.J., and Hwang, C.B.C. 1999. Effects of exonuclease activity and nucleotide selectivity of the herpes simplex virus DNA polymerase on the fidelity of DNA replication in vivo. *J Virol* 73:5326–5332.

Hwang, Y.T., Liu, B.-Y., and Hwang, C.B.C. 2002. Replication fidelity of the *supF* gene integrated in the thymidine kinase locus of herpes simplex virus type 1. *J Virol* 76:3605–3614.

Ikehata, H., Akagi, T., Kimura, H., Akasaka, S., and Kato, T. 1989. Spectrum of spontaneous mutations in a cDNA of the human *hprt* gene integrated in chromosomal DNA. *Mol Gen Genet* 219:349–358.

Kroutil, L.C., Frey, M.W., Kunkel, T.A., and Benkovic, S.J. 1998. Effect of accessory proteins on T4 DNA polymerase replication fidelity. *J Mol Biol* 278:135–146.

Kunkel, T.A. 1985. The mutational specificity of DNA polymerase-β during *in vitro* DNA synthesis. *J Biol Chem* 260:5787–5796.

Kunz, B.A., Kohalmi, L., Kang, X.L., and Magnusson, K.A. 1990. Specificity of the mutator effect caused by disruption of the *RAD1* excision repair gene of *Saccharomyces cerevisiae*. *J Bacteriol* 172:3009–3014.

Lee, G.S.-F., Savage, E.A., Ritzel, R.G., and von Borstel, R.C. 1988. The base-alteration spectrum of spontaneous and ultraviolet radiation-induced forward mutations in the *URA3* locus of *Saccharomyces cerevisiae*. *Mol Gen Genet* 214:396–404.

Levitt, S.D. and Dubner, S.J. 2006. *Freakonomics. A Rogue Economist Explores the Hidden Side of Everything* (revised and expanded). New York: HarperCollins.

Lichtenauer-Kaligis, E.G.R., Thijssen, J., den Dulk, H., van de Putte, P., Tasseron-de Jong, J.G., and Giphart-Gassler, M. 1993. Genome wide spontaneous mutation in human cells determined by the spectrum of mutations in hprt cDNA genes. *Mutagenesis* 8:207–220.

Little, J.W. 1991. Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochemie* 73:411–422.

Loeb, L.A., Loeb, K.R., and Anderson, J.P. 2003. Multiple mutations and cancer. *Proc Natl Acad Sci USA* 100:776–781.

Lu, O., Hwang, Y.T., and Hwang, C.B.C. 2002. Mutation spectra of herpes simplex virus type 1 thymidine kinase mutants. *J Virol* 76:5822–5828.

Mackwan, R.R., Carver, G.T., Drake, J.W., and Grogan, D.W. 2007. An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloferax volcanii*. *Genetics* 176:697–702.

Maitra, M., Gudzelak, Jr. A., Li, S.-X., Matsumoto, Y., Eckert, K.A., Jager, J., and Sweasy, J.B. 2002. Threonine 79 is a hinge residue that governs the fidelity of DNA polymerase b by helping to position the DNA within the active site. *J Biol Chem* 277:35550–35560.

Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W., and García-Arenal, F. 2002. The rate and character of spontaneous mutation in an RNA virus: unusual pattern and multiple mutations. *Genetics* 162:1505–1511.

McCool, J.D., Long, E., Petrosino, J.F., Sandler, H.A., Rosenberg, S.A., and Sandler, S.J. 2004. Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. *Mol Microbiol* 53:1343–1357.

Mo, J.-Y., Maki, H., and Sekiguchi, M. 1991. Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. *J Mol Biol* 222:925–936.

Nick McElhinny, S., Stith, C.M., Burgers, P.M.J., and Kunkel, T.A. 2007. Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of *Saccharomyces cerevisiae* DNA polymerase δ. *J Biol Chem* 282:2324–2332.

Ninio, J. 1991. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics* 129:957–962.

Ninio, J. 1996. Gene conversion as a focusing mechanism for correlated mutations: a hypothesis. *Mol Gen Genet* 251:503–508.

Ninio, J. 2000. Illusory defects and mismatches: why must DNA repair always be (slightly) error prone?. *BioEssays* 22:396–401.

Oller, A.R. and Schaaper, R.M. 1994. Spontaneous mutation in *Escherichia coli* containing the *dnaE911* DNA polymerase antimutator allele. *Genetics* 138:263–270.

Opresko, P.L., Sweasy, J.B., and Eckert, K.A. 1998. The mutator form of polymerase β with amino acid substitution at tyrosine 265 in the hinge region displays an increase in both base substitution and frame shift errors. *Biochemistry* 37:2111–2119.

Osheroff, W.P., Jung, H.K., Beard, W.A., Wilson, S.H., and Kunkel, T.A. 1999. The fidelity of DNA polymerase β during distributive and processive DNA synthesis. *J Biol Chem* 274:3642– 3650.

Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. 2005. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol Cell* 19:791–804.

Romac, S., Leong, P., Sockett, H., and Hutchinson, F. 1989. DNA base changes induced by ultraviolet light mutagenesis of a gene on a chromosome in Chinese hamster ovary cells. *J Mol Biol* 209:195–204.

Rossi, A.M., Thijssen, J.C., Tates, A.D., Vrieling, H., Natarajan, A.T., Lohman, P.H., and van Zeeland, A.A. 1990. Mutations affecting RNA splicing in man are detected more frequently in somatic than in germ cells. *Mutat Res* 244:353–357.

Sargentini, N.J. and Smith, K.C. 1994. DNA sequence analysis of gamma-radiation (anoxic)-induced and spontaneous *lacI*^{d} mutations in *Escherichia coli* K-12. *Mutat Res* 309:147–163.

Schaaper, R.M. 1988. Mechanisms of mutagenesis in the *Escherichia coli* mutator *mutD5*: role of DNA mismatch repair. *Proc Natl Acad Sci USA* 85:8126–8130.

Schaaper, R.M. 1993. The mutational specificity of two *Escherichia coli dnaE* antimutator alleles as determined from *lacI* mutation spectra. *Genetics* 134:1031–1038.

Schaaper, R.M. and Dunn, R.L. 1987. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc Natl Acad Sci USA* 84:6220–6224.

Schaaper, R.M. and Dunn, R.L. 1991. Spontaneous mutation in the *Escherichia coli lacI* gene. *Genetics* 129:317–326.

Schaaper, R.M., Danforth, B.N., and Glickman, B.W. 1986. Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J Mol Biol* 189:273–284.

Tindall, K.R. and Stankowski, L.F. 1989. Molecular analysis of spontaneous mutations at the *gpt* locus in Chinese hamster ovary (AS52) cells. *Mutat Res* 220:241–253.

Venkatesan, R.N., Hsu, J.J., Lawrence, N.A., Preston, B.D., and Loeb, L.A. 2006. Mutator phenotypes caused by substitution at a conserved motif A in eukaryotic DNA polymerase δ. *J Biol Chem* 281:4486–4494.

Wang, F.J. and Ripley, L.S. 1998. The spectrum of acridine resistant mutants of bacteriophage T4 reveals cryptic effects of the *tsL141* DNA polymerase allele on spontaneous mutagenesis. *Genetics* 148:1655–1665.

Wang, J., Gonzalez, K.D., Scaringe, W.A., Tsai, K., Li, N., Gu, D., Li, W., Hill, K.A., and Sommer, S.S. 2007. Evidence for mutation showers. *Proc Natl Acad Sci USA* 104:8403–8408.

Watson, D.E., Cunningham, M.L., and Tindall, K.R. 1988. Spontaneous and ENU-induced mutation spectra at the *cII* locus in Big Blue® Rat2 embryonic fibroblasts. *Mutagenesis* 13:487–497.

Witkin, E.M. and Parisi, E.C. 1974. Bromouracil mutagenesis: mispairing or misrepair?. *Mutat Res* 25:407–409.

Witkin, E.M. and Sicurella, N.A. 1964. Pure clones of lactose-negative mutants obtained in *Escherichia coli* after treatment with 5-bromouracil. *J Mol Biol* 8:610–613.

Editor: Susan M. Rosenberg